International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-9, Issue-10; Oct, 2023
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.910.2

Deep Learning Neural Networks in the Cloud

Burhan Humayun Awan

burhanhumayunawan@yahoo.ie

Received: 05 Sep 2023; Received in revised form: 08 Oct 2023; Accepted: 15 Oct 2023; Available online: 25 Oct 2023

***Abstract*— *Deep Neural Networks (DNNs) are currently used in a wide range of critical real-world applications as machine learning technology. Due to the high number of parameters that make up DNNs, learning and prediction tasks require millions of floating-point operations (FLOPs). Implementing DNNs into a cloud computing system with centralized servers and data storage sub-systems equipped with high-speed and high-performance computing capabilities is a more effective strategy. This research presents an updated analysis of the most recent DNNs used in cloud computing. It highlights the necessity of cloud computing while presenting and debating numerous DNN complexity issues related to various architectures. Additionally, it goes into their intricacies and offers a thorough analysis of several cloud computing platforms for DNN deployment. Additionally, it examines the DNN applications that are already running on cloud computing platforms to highlight the advantages of using cloud computing for DNNs. The study highlights the difficulties associated with implementing DNNs in cloud computing systems and provides suggestions for improving both current and future deployments.*

***Keywords*— Deep Learning, Neural Networks, Cloud Computing.**

I. INTRODUCTION

Numerous pattern recognition applications in real-world fields like e-commerce, manufacturing, medicine and health, and autonomous vehicles are now being developed using deep neural networks (DNNs). However, due to their extensive parameter requirements, DNNs pose significant computing demands, especially during training. DNNs typically have millions of parameters. As an illustration, popular DNNs like AlexNet have 60 million parameters, while VGG-16 has 138 million. A DNN with 175 billion parameters that required a lengthy seven months to train was used in a recent OpenAI project for natural language processing (NLP) [1]. As a result, it is impossible to train a large DNN using a single isolated computer. High-performance computing tools are necessary for the efficient training of DNNs.

DNN deployment on cloud platforms has increased in popularity recently. These cloud computing platforms are extremely fast and memory-capable high-performance computing systems. On a variety of cloud machine learning (ML) platforms, such as Google Colab and Amazon Web Services (AWS) Deep Learning, training can be effectively carried out in reasonable amounts of time. Centralized servers powered by cloud computing provide a lot of computing power, a lot of data storage, fast processing, low latency, and high availability. DNNs for online applications can be deployed thanks in large part to cloud computing.

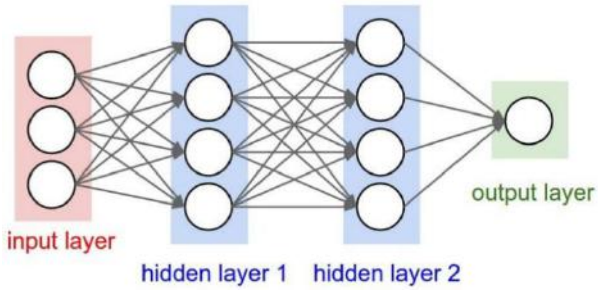

Fig.1: Deep Neural Network with hidden layers

The implementation of cloud platforms for computationally intensive tasks is discussed in various recent survey publications cited in the current manuscript. These survey papers are divided into three main categories: applications, performance improvement technologies, and security technologies. Yan et al.'s [2] study of security technologies covered a range of solutions for preventing harmful assaults, including talks of countermeasures. Recent

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

security methods, including encryption, access structure, and fine-grained trace mechanisms, were introduced by Nita and Mihailescu [3] and Sun et al. [4]. Gai et al.'s [5] analysis of the functionality of blockchain-enabled integrated hardware and software in cloud data centers was centered on the integration of blockchain with clouds.

Regarding performance improvement, Pupykina et al. [7] investigated memory management methods in cloud computing, and Xu et al. [6] assessed technologies for managing virtual machine performance. Offloading technologies for job optimization across cloud and edge systems were covered by Wang et al. [8]. In order to address problems and look toward the future, Xu et al. [6] provide an overview of computational distribution strategies for managing virtual machines in the cloud. Deep reinforcement learning-based cloud resource scheduling solutions were discussed by Zhou et al. [9] and Feng et al. [10].

Regarding applications, Bera et al. [12] reviewed cloud computing applications in smart grids, while Khan et al. [11] provided an overview of mobile cloud topologies and the advantages of cloud computing. An overview of cloud computing architectures for cyber-physical systems was presented by Cao et al. [13], evaluating numerous applications. Notably, the development of DNNs in cloud computing systems is not a specific emphasis of these survey studies.

This research intends to close this gap by offering a thorough analysis of cloud computing methods for DNN deployment, along with considerations of difficulties and potential future research areas. The article's remaining sections are structured to present different DNN mechanisms, the need for cloud computing, popular cloud platforms for deploying DNNs, specific DNN applications implemented in cloud systems, difficulties in current DNN deployments using cloud computing systems, and opportunities for improving current DNN deployments on cloud systems. Finally, a summary of the findings in brings the article to a close.

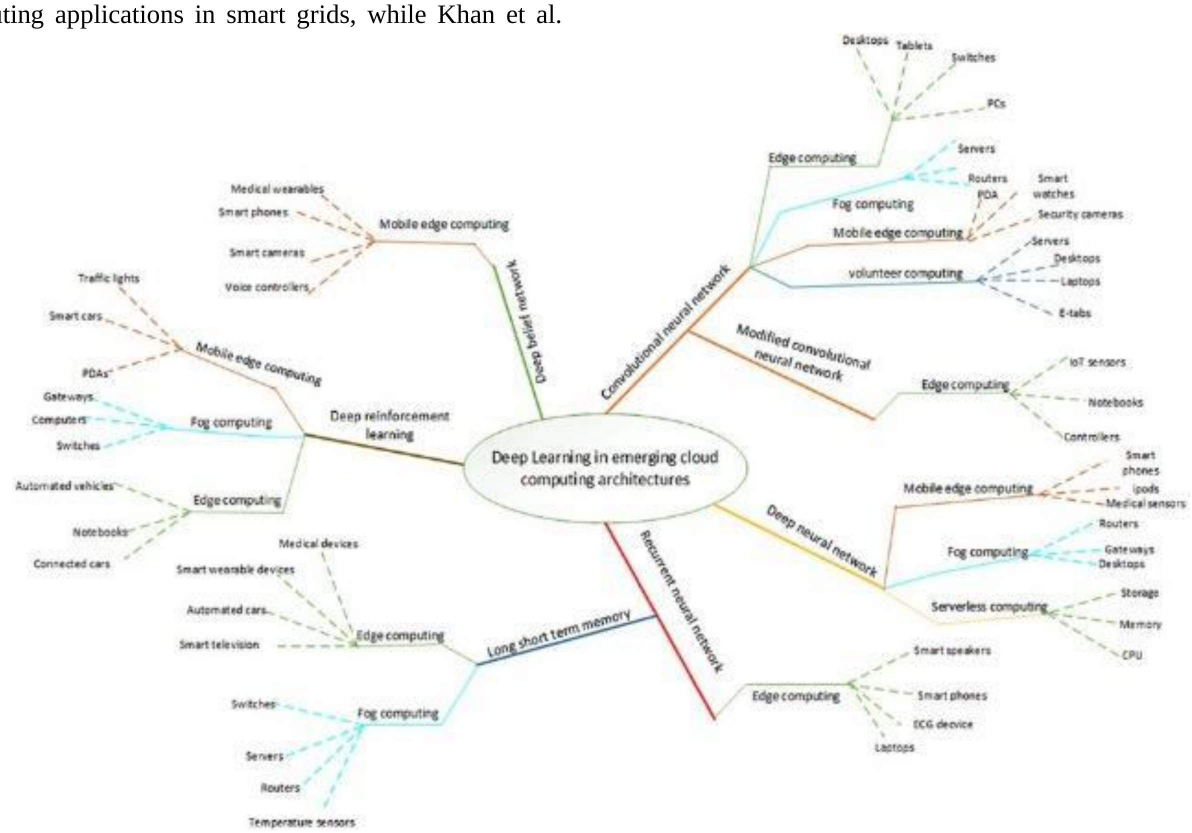

Fig.2: Deep learning architectures in emerging cloud computing architectures

(Source: Fatsuma Jauro, 2020. Deep learning architectures in emerging cloud computing architectures. Volume 96, 106582, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2020.106582)

Computational Complexities in Deep Learning

A single deep neural network (DNN) consists of a significant number of parameters, demanding a large amount of storage memory. A DNN's training and execution phases both need a considerable time commitment. This section explores several popular DNN architectures, such as multilayer perceptron's (MLPs), convolutional neural networks (CNNs), and graph neural networks (GNNs), all

This article can be downloaded from here: www.ijaems.com

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License. http://creativecommons.org/licenses/by/4.0/

of which have intricate architectural designs and a wide range of DNN parameters. Because DNN training requires extensive computation, it is impractical to use a single isolated computer for this purpose, which highlights the vital role that cloud computing plays in aiding DNN training.

An extensively discussed neural network is the multilayer perceptron [24–26]. Each of MLP's levels, which include an input layer, hidden layers, and an output layer, has a collection of perceptron components, also known as neurons. An MLP with two hidden layers, an input layer, and an output layer is shown in Figure 3.

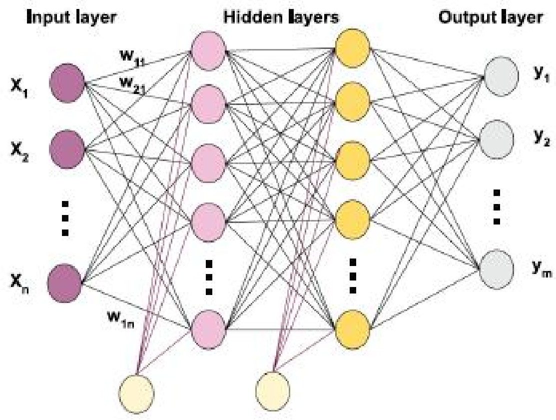

Fig.3: Multi-Layer Perceptron's Typical Topology

Longer computational times are required to optimize an MLP when number of layers are higher. Multiple apps have been created as a result of using MLP in the cloud. For instance, the study described in [28] concentrated on developing a forecasting model that made use of multiple input variables generated from several daily basic food price kinds. Using the Amazon Cloud Services infrastructure, this model sought to forecast the Surabaya consumer price index. The multilayer perceptron technique was used in the study to build a prediction system with a hidden layer, an epoch, and a set number of neurons. Similar to this, a different study [29] divided regions afflicted by cancer and stored relevant information in the cloud using transfer learning-based cancer segmentation (TL-CAN-Seg) technology. A unique MLP and an altered Levenberg-Marquardt (LM) algorithm were used to interpret complex picture patterns and accomplish precise classification of areas affected by breast cancer, improving the accuracy of breast cancer diagnosis.

Recurrent neural networks (RNNs) are more effective than MLPs at handling temporal data, including text and sequentially correlated time series. The result from the previous phase in the sequence serves as the input for the next step in RNNs' unique versions of neural processing units [30]. RNNs' hidden state, which is used for iterative processing, catches and holds onto data all the way through the sequence. RNNs have the ability to learn by storing, retrieving, and using historical data to make predictions. Model interactions at different temporal scales endow memory. Information from all earlier steps is captured by the aforementioned concealed state. In order to produce the output at a particular step in the sequence, the trained RNN can combine the input sequence and the hidden state. Despite being created more than a decade ago, RNNs still have issues with memory storage and computational time limits [31].

Figure 4 depicts a typical CNN with input/output layers, hidden layers, and a fully connected network [41,42]. Convolution, activation, and pooling layers are all included in a hidden layer. The activation function assists in learning nonlinear input patterns from the convolution output, the pooling layer consolidates the outputs of the activation function into a single value, and the convolution layer retrieves input features within this layer. To make categorization easier, relevant features are retrieved after numerous convolution and pooling procedures.

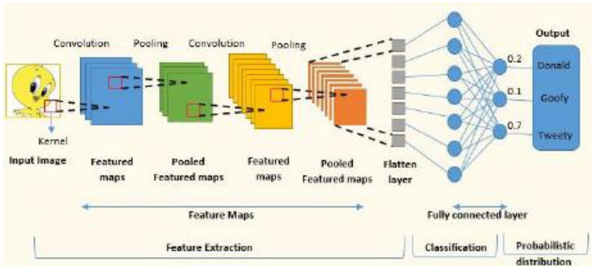

Fig.4: Typical Convolutional Neural Network (CNN)

(Source:https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/)

A CNN typically consists of millions of network parameters, each of which must be determined over the course of a long period of time. In particular, ShuffleNet, GoogLeNet, DenseNet, ResNet, AlexNet, VGGNet, and ConvNet are a few CNN models that include millions of network parameters and need a lot of training time. For instance, it takes more than two weeks to train ConvNet on a machine with four NVIDIA Titan Black GPUs. Additionally, OpenAI's CNN, also known as neural architecture search (NAS), requires six months to train when using 8 P100s in parallel scaling and has a remarkable 175 billion parameters for natural language processing.

While purchasing many computers for a single CNN proves economically inefficient, training a big CNN on a sole computer is prohibitively time-consuming and unfeasible. Therefore, the use of CNNs in cloud infrastructure has been proven in a number of applications [49-56].

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

GNN is a result of graph representation learning [57–59], which entails transforming and comprehending nodes and edges in a graph into a continuous space with fewer dimensions. GNN treats non-Euclidean domains using elaborate data structures that reflect the relationships between these entities [60]. This is true even if it works with Euclidean 1-D sequences like texts and 2-D grids like images [61]. In GNNs, the underlying data structure is represented by a graph with the formula G = (V, E), where V denotes the set of vertices or nodes and E denotes the edges joining them. It is possible for the relationship (u, v) E to be symmetric or asymmetric. Graphs can be homogenous, like social networks like Facebook friends, or diverse, like knowledge graphs. The incorporated elements or graphic topologies may also alter dynamically over time.

GNNs' exponential node expansion increases computational complexity and memory use significantly [62]. Notably, some sizable GNNs, like GemNet-XL, include billions of parameters [69]. Existing GNN models are judged insufficient for large-scale graphs that incorporate intricate topologies because they have only been evaluated on small graphs [70]. Examples of GNN implementation in cloud infrastructure have shown that graph analysis is scalable and effective in a variety of applications, such as recommender systems, traffic flow prediction, industrial IoT, privacy preservation, and matrix completion [71–76].

It is essential to comprehend the causes of latency in the training and inference phases of various DNN architectures if you want to deploy them in the cloud with the least amount of delay possible. Using methods like dilated convolutions in CNNs, for instance, one can increase the network's receptive field without adding more parameters or layers, which lowers computational cost and inference delay [77,78]. The use of backpropagation gradients through time, which can be computationally expensive, is not necessary with randomization-based learning techniques like echo state networks (ESN) [79]. Additionally, GNN and CNN pruning strategies can reduce the amount of parameters and computations needed, resulting in quicker inference times [80,81].

Cloud computing architectures for deep learning based applications

Because deep neural network (DNN) architectures are complex and demand a large number of parameters, training and execution periods are prolonged. As a result, it is impossible to train or deploy DNN using a single standalone computer. Cloud computing offers a practical answer to this problem for these kinds of resource-intensive computations. Cloud computing meets the demanding computational needs of several DNN implementations and training tasks by offering substantial computing power and abundant data storage, eventually helping customers using DNNs in intense applications [82].

The structure and composition of cloud data centers are outlined in the next section. The next part discusses frequently used commercial cloud computing platforms for the deployment of DNN and provides an overview of public or volunteer cloud computing platforms. The section also discusses frequently used cloud streaming systems, offering light on how data streaming is implemented there. Researchers studying deep learning (DL) who are looking for reliable, affordable, and quick computing platforms for DNN development will find this comprehensive material to be of particular value.

- Cloud data centers

Data storage and computing are handled by cloud data centers or remote clouds, including backhaul and core networks [83]. A typical cloud computing architecture, shown in Figure 5, is made up of cloud users, internet network providers, and cloud service providers. Users provide computational data over network service providers, which servers then receive. This data is processed using cloud resources, ensuring that users have enough access to a shared pool of resources in response to their requests. This use of cloud resources makes it possible to offer adaptable processing power and storage, which eventually helps well-known cloud-based companies like Amazon and Google Cloud stay profitable [84, 85].

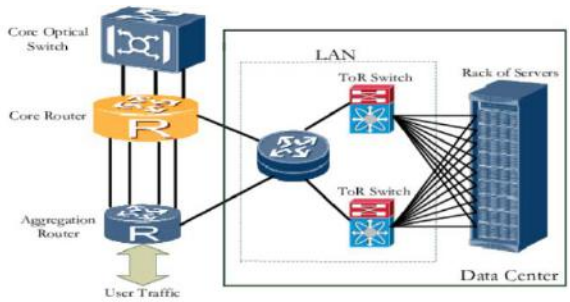

Fig.5: Typical Cloud Data Center Architecture

(Source: Elmirghani. 2018. GreenTouch GreenMeter Core Network Energy-Efficiency Improvement Measures and Optimization. Journal of Optical Communications and Networking. 10. 10.1364/JOCN.10.00A250)

Users' requests for calculations are spread across a variety of cloud platforms with numerous data centers [86]. Resource-intensive computations are made possible by resource sharing within and between data centers. To increase processing power, a distributed cloud can also be connected to hybrid clouds, public clouds, and edge

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

computing installations. User requests can be distributed among close-by data centers in an effort to reduce data transmission delay. The architecture of a distributed cloud consists of numerous sub-clouds. Based on the availability of resources, a central controller within the distributed cloud distributes computing workloads among various sub-clouds.

- DL in the cloud

The availability of customizable computing resources, a critical component for a variety of DL workloads characterized by different degrees of compute needs across distinct activities and datasets, is one of the main advantages of cloud computing. Numerous cloud providers offer services to meet these needs, such as the auto-scaling features of Amazon EC2 and the scale-up and scale-out capabilities of Microsoft Azure. These services make it possible to run DL workloads efficiently even while working with limited cloud resources.

Parameter server

To enable the scalability of distributed machine learning applications within cloud data centers, parameter servers (PSs) have been developed [87]. To train deep neural networks, PSs have been incorporated into a number of deep learning frameworks, including TensorFlow and MXNet. Even with developments, total reliability cannot be assured because cloud data centers may have server outages. To ensure the orderly completion of learning tasks in such contexts, preventative procedures are required to handle job-sharing and backups [88]. The collective, which consists of a server group and a worker group, is the center of the PS framework. Both have job schedulers and worker nodes, which work together in DNN training, while the latter also includes a manager and server nodes. Consistent hashing is used to express shared parameters as vectors of (key, value). Operational data created by nodes is communicated to the server, which then distributes global information to each node. This framework provides flexibility in guaranteeing consistent data when the algorithm is not sensitive to data inconsistencies, increasing the reliability of the PS framework. It allows asynchronous tasks and dependencies by initiating the necessary methods.

Due to its widespread connectivity with a sizable and dynamic resource pool, the PS architecture is more suitable with heterogeneous production data centers and public clouds than alternative techniques like AllReduce5. But the original PS architecture has some drawbacks, such as imbalances, elasticity constraints, and static parameter assignments. It is not possible to incorporate more available resources into ongoing training activities, and workload distribution across nodes frequently does not maximize resource capacity. Several strategies have been put out to deal with these limitations and improve cloud computing's capabilities. For instance, Proteus, an elastic PS framework created to scale up training on public clouds, was introduced by Harlap et al. [89]. The framework dynamically assigns PSs and personnel using three transitional stages, maximizing cost reductions, particularly when temporary revocable resources become available. Litz [90] created logical executors, which map physical nodes to control the executor states of specific applications, in an additional effort to increase elasticity. This method also includes micro-tasks for determining dependencies and assigning micro-tasks accordingly.

Advanced learning frameworks

Deep neural network (DNN) training in cloud environments will be more dependable and effective thanks to proposed improvements in learning frameworks. For workload scheduling and dynamic resource scaling, the DL-driven framework DL2 [91] combines supervised learning and reinforcement learning techniques. The DNN is initially taught offline to assimilate resource allocation patterns from prior judgments, and reinforcement learning is then used to improve the training of the DNN. An exploitation-exploration method was used by Chen et al. [92] to propose the dynamic PS load distribution scheme known as PSLD. The plan consists of three stages: gathering data on each PS, having workers create performance profiles, and choosing PSs based on performance indicators and communication time.

At the same time, Wang et al. [93] suggested the elastic parameter server (EPS), a simple solution that enables dynamic resource allocation and deallocation for increased resource use and training efficiency. To improve scalability and optimize resource utilization, this strategy incorporates heuristic scheduling modes like incoming job scheduling and running job scheduling. Additionally, more specialized frameworks have been created, with a particular emphasis on DL workloads on private clouds. In order to reduce waiting times, Hu et al. [94 - 95] implemented an optimization technique based on the idea of training progress and integer programming to handle resource scaling challenges particular to AWS and Huawei clouds.

- Data streaming for the cloud

The continuous streaming of real-time data is essential for supporting deep learning (DL) applications in various sensor networks and control systems, such as those found in autonomous vehicles or smart grids. For activities like recognition or decision-making, it is essential to acquire measurements or data as soon as possible. Lack of access to the most recent data versions may jeopardize system functionality and lower safety standards. Given the huge amounts of data volumes contained in each sample, it is

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

impractical to retrain deep neural networks (DNNs) using a standalone computer, highlighting the necessity of data streaming in a cloud setting.

Cloud data streaming platforms

Platforms for streaming cloud data have found use in government and academic settings for the analysis of data gathered by sensor networks. To make geographic data processing easier, the Southern California Earthquake Center, for instance, has set up a geophysical sensor network [96]. This network's tens of thousands of sensors continuously and quickly sample data. In order to better understand climate change and create systems for predicting earthquakes and inland flooding, geospatial data must be collected. Furthermore, to study seismic and hydrological features in North America, the Geodesy Advancing Geosciences and EarthScope (GAGE) GPS network uses information from more than a thousand GPS sensors [97]. Similar to this, the US National Science Foundation has provided funding for the creation of a worldwide sensor network that will largely be used to study climate change and the cycling of carbon [98].

Data-streaming approaches

Given that many deep neural networks (DNNs) operate in dynamic contexts, it is essential that they constantly absorb new information or undergo retraining. A data-streaming technique has been developed to handle this issue and determine whether streaming data is required for changing DNN parameters [101]. This method incorporates a strategic trade-off between training expenses and performance to decide if DNNs need to be updated. Its use with TensorFlowOnSpark for three online learning workloads has shown a decrease in total processing time. Similar to this, Ashfahani et al. [102] have suggested a data-streaming approach to modify network topologies in response to fresh input. This method makes it easier for network nodes to grow and shrink dynamically, improving performance while reducing complexity. In comparison to previous approaches, comparative evaluations on standard datasets have shown improved network performance and decreased network complexity [102].

A similar advancement is the introduction of an incremental high-order DL model by Li et al. [103] that is designed to adapt to high-frequency data streams. The strategy efficiently minimizes adaptation time by translating data into a high-order tensor space, and then uses first-order approximations to reduce the time-consuming parameter incrementation frequently associated with iterative procedures. DNNs may now adapt to dynamic situations better than traditional iterative approaches, effectively satisfying real-time needs. A unique fuzzy neural network has been introduced by Pratama et al. [104] that automatically incorporates fuzzy rules from data streams, using a simplification procedure to merge unnecessary hidden layers and control network growth.

Results from experiments show that this approach successfully limits network size while maintaining performance standards. In addition, Nguyen et al. [105] have developed a sensor network for gathering maritime data, using a deep recurrent neural network combined with streaming data to monitor fishing activities, spot smuggling, forecast maritime pollution, and improve maritime traffic safety and security in real time. This comprehensive method effectively handles noisy and infrequently sampled data in maritime environments by combining latent variable modeling and data streaming to capture key elements within maritime dynamics.

In order to provide real-time predictions, DNN designs in cloud environments must incorporate minimal inference latencies. Smaller DNN designs or the use of accelerators, together with the storage of data features in low-latency storage locations for offline precomputing predictions, can improve serving latency. Additionally, essential for adjusting to newly streamed data and improving DNN performance over time is the implementation of incremental training [106]. Using model artifacts from well-known, publicly accessible DNNs, this entails routine updates of DNNs based on fresh streams of data, enabling updates without the requirement for retraining from scratch.

II. APPLICATIONS OF DNNS IN THE CLOUD

Deep neural network (DNN) deployment in the cloud has become widely used in a wide range of applications. The sections that follow provide an overview of a few of these applications, including wireless capsule endoscopy, travel, natural language processing (NLP), business intelligence (BI), cybersecurity, anomaly detection, and mobile-cloud-assisted implementations. Tables that give a brief synopsis of the application's content follow these subsections. These descriptions accurately identify the complex research issues posed by the combination of cloud systems and DNNs.

- Natural language processing

The application of artificial intelligence (AI) to the interplay of computers and human language is known as natural language processing (NLP). It entails the creation of computational models and algorithms that provide computers the ability to effectively comprehend, decode, and produce human language. NLP includes a wide range of activities, including sentiment analysis, speech recognition, language translation, natural language comprehension, and more. Its uses span from text analysis

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

tools and chatbots to virtual assistants and language translation programs [107 - 115].

- Business intelligence

The technology, systems, and procedures utilized for the gathering, integrating, analyzing, and presenting of business information are referred to as business intelligence (BI). Utilizing a range of software tools, it entails gathering information from internal and external sources and turning it into insights that can be used to inform strategic and tactical business choices. Data mining, data analysis, querying, and reporting are just a few of the tasks included in business intelligence (BI), which frequently uses metrics and key performance indicators (KPIs) to assess an organization's or business's success. By offering historical, current, and predictive insights of corporate operations, BI's primary goal is to enable better business decision-making [117 – 120].

- Cybersecurity

The practice of defending computer systems, networks, programs, and data from online attacks, harm, or illegal access is known as cybersecurity. In order to protect information technology (IT) systems and networks from theft, damage, or disruption while maintaining the integrity, confidentiality, and availability of data, security policies and safeguards must be put in place. Malware, ransomware, phishing scams, hacking, and other types of cybercrime are just a few of the risks that cybersecurity seeks to protect against. It includes a wide range of tactics, tools, and best practices for preventing, spotting, and dealing with digital threats and security breaches [121 - 135].

- Anomaly detection

Finding patterns or occurrences within a dataset that drastically depart from expected behavior is known as anomaly detection. Finding anomalies, outliers, or abnormalities that deviate from the predicted patterns or trends in the data entails using statistical analysis, machine learning methods, and data mining approaches. The identification of anomalous or suspicious activity is critical for assuring the integrity, security, and effectiveness of operations in a number of different sectors, including cybersecurity, fraud detection, system health monitoring, and industrial quality control [140 – 154].

- Travel

AI efficiently organizes and processes large datasets produced by both clients and service providers in the tourist and hospitality industries. Notably, a sizeable portion of pertinent data used in the tourism sector comes from GPS applications and is frequently linked to social media, Internet of Things (IoT), and site traffic statistics. These huge databases are managed within the framework of the "smart" tourism industry, with the goal of providing passengers with knowledgeable and personalized services. Applying sophisticated intelligent approaches for analysis is necessary when working with datasets that are so diverse, detailed, and dispersed [157 – 176].

- Remote medical diagnosis

The use of wireless capsule endoscopy (WCE) has significantly increased over the past two years. These methods provide a level of internal human visibility for diagnostic purposes that is comparable to standard endoscopy. These technologies were initially presented in 2000, and after going through clinical studies, they were given the go-ahead by the Food and Drug Administration in 2001. Notably, these technologies provide improved portability and have numerous uses in the delivery of systemic biologics and healthcare services [177 – 190].

- Mobile-cloud-assisted applications

A developing technology with a wide range of possible applications, the idea of mobile cloud-assisted applications. The ability to transfer tasks to cloud servers, hence prolonging the system's operational lifespan, is the main goal of mobile cloud computing (MCC). The computing strain on mobile devices like smartphones, tablets, and iPads is also lessened by MCC. To ensure the effectiveness of the job offloading inference engine and to ease resource restrictions on smartphones, which often have much less processing power compared to older approaches, an extensive evaluation based on traced data is carried out [195 – 197].

Before using Deep Neural Networks (DNNs), it is urgently necessary to remove unnecessary and redundant frames during the Wireless Capsule Endoscopy (WCE) operation in order to prioritize the video content. However, there are some significant difficulties with video prioritizing in WCE, especially when there are limited resources and processing capabilities. Because of this, the integration of MCC helps to offer affordable storage, robust computational power, and software services [198 – 199].

III. CHALLENGES AND FUTURE DIRECTIONS

Cloud-based Deep Neural Networks (DNNs) have been constructed for a variety of applications requiring extensive big data analysis and high-performance computing, making use of the significant computational power and data storage capacities provided by cloud platforms. Despite the fact that DNNs are effective tools for pattern recognition, they pose a number of research obstacles, including issues with energy consumption, the length of training and execution, data security, and cloud compatibility [203 – 275].

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

- Energy efficiency
- Training cost
- Scalability
- Data security
- Privacy awareness
- Cloud interoperability
- Learning from non-stationary data: retraining efficiency and adaptation
- Elastic implementations of deep learning models and flexible resource allocation
- Deep reinforcement learning
- 3D vision applications
- Optimization of DNNs

IV. CONCLUSIONS

DNNs have a wide range of applications in numerous real-world fields. Deploying DNNs on solitary stand-alone PCs or mobile devices, however, is frequently problematic for huge data storage and analysis applications due to their computational complexity and the enormous number of parameters necessary in training. As a result, the use of DNNs in cloud computing systems has attracted a lot of interest. In the beginning, this review article lays out the justification for using and training DNNs in cloud-based systems. It then digs into the intricate computations of popular DNNs, such as MLP, CNN, RNN, and GNN, highlighting their high parameter and FLOPs needs.

The research also provides a thorough analysis of volunteer and public cloud computing platforms that have effectively included and applied DNNs. Researchers and software developers can use this information to choose the best cloud computing platform for their DNN-focused apps. The paper also provides information on a number of application fields, including NLP, BI, cybersecurity, anomaly detection, and travel, which have recently reaped significant advantages from the integration of DNN in cloud computing. It outlines the key difficulties involved in this approach while highlighting the benefits and efficiency of installing DNNs in cloud-based applications.

The research also suggests potential possibilities to improve current deployments using cloud computing platforms and DNNs. This thorough overview study is projected to be a useful resource for researchers and programmers interested in successfully implementing DNNs on cloud computing platforms.

REFERENCES

[1] E. Strubell, A. Ganesh, and A. McCallum, Energy and policy considerations for deep learning in nlp, in: Annual Meeting of the Association for Computational Linguistics, 2019.

[2] Q. Yan, F.R. Yu, Q.X. Gong, J.Q. Li, Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: a survey, some research issues, and challenges, IEEE Communications Surveys and Tutorials 18 (1) (2016) 602–622.

[3] S.L. Nita and M.I. Mihailescu, On artificial neural network used in cloud computing security – a survey, in Proceedings of the International Conference of Electronics, Computers and Artificial Intelligence, 2018.

[4] P.J. Sun, Privacy protection and data security in cloud computing: a survey, challenges, and solutions, IEEE Access, vol. 7, pp. 147 420–147 452, 2019.

[5] K. Gai, J. Guo, L.H. Zhu, S. Yu, Blockchain meets cloud computing: a survey, IEEE Transactions on Neural Networks and Learning Systems 22 (3) (2022) 2009–2030.

[6] F. Xu, F.M. Liu, H. Jin, and A.V. Vasilakos, Managing performance overhead of virtual machines in cloud computing: A survey, state of the art, and future directions, Proceedings of the IEEE, vol. 102, no. 1, pp. 11–31, 2014.

[7] A. Pupykina and G. Agosta, Survey of memory management techniques for HPC and cloud computing, IEEE Access, vol. 7, pp. 167 351–167 373, 2019.

[8] B. Wang, C.C. Wang, W.W. Huang, Y. Song, X.Y. Qin, A survey and taxonomy on task offloading for edge-cloud computing, IEEE Access 8 (2020) 186 080–186 101.

[9] G. Zhou, W. Tian, R. Buyya, Deep reinforcement learning-based methods for resource scheduling in cloud computing: A review and future directions, Journal of Cloud Computing 11 (2022), paper number 3.

[10] Y. Feng and F. Liu, Resource management in cloud computing using deep reinforcement learning: A survey, in Proceedings of the 10th Chinese Society of Aeronautics and Astronautics Youth Forum, 2023, pp. 635–643.

[11] A.R. Khan, M. Othman, S.A. Madani, S.U. Khan, A survey of mobile cloud computing application models, IEEE Communications Surveys and Tutorials 16 (1) (2014) 393–413.

[12] S. Bera, S. Misra, J.J.P.C. Rodrigues, Cloud computing applications for smart grid: a survey, IEEE Transactions on Parallel and Distributed Systems 26 (5) (2015) 1477–1494.

[13] K. Cao, S.Y. Hu, Y. Shi, A.W. Colombo, S. Karnouskos, X. Li, A survey on edge and edge-cloud computing assisted cyber-physical systems, IEEE Transactions on Industrial Informatics 17 (11) (2021) 7806–7819.

[14] D. Soni, N. Kumar, Machine learning techniques in emerging cloud computing integrated paradigms: A survey and taxonomy, Journal of Network and Computer Applications 205 (2022).

[15] T. Khana, W.H. Tiana, R. Buyya, Machine learning (ml)-centric resource management in cloud computing: A review and future directions, Journal of Network and Computer Applications 204 (2022).

[16] A. Saiyeda, M.A. Mir, Cloud computing for deep learning analytics: a survey of current trends and challenges, International Journal of Advanced Research in Computer Science 8 (2) (2017).

[17] P.S. Priya, P. Malik, A. Mehbodniya, V. Chaudhary, A. Sharma, and S. Ray, The relationship between cloud

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

computing and deep learning towards organizational commitment, in: Proceedings of the 2nd International Conference on Innovative Practices in Technology and Management, 2022.
[18] F. Benedetto and A. Tedeschi, Big data sentiment analysis for brand monitoring in social media streams by cloud computing, in: Sentiment Analysis and Ontology Engineering. Springer, 2016, pp. 341–377.
[19] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D.C. Anastasiu, Stock price prediction using news sentiment analysis, in: 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2019, pp. 205–208.
[20] S. Prasomphan, Improvement of chatbot in trading system for smes by using deep neural network, in: 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, 2019, pp. 517–522.
[21] D. Scheinert, A. Acker, L. Thamsen, M.K. Geldenhuys, and O. Kao, Learning dependencies in distributed cloud applications to identify and localize anomalies, in: Proceedings of IEEE/ACM International Workshop on Cloud Intelligence, 2021, pp. 7–12.
[22] M.A. Elsayed, M. Zulkernine, Predictdeep: Security analytics as a service for anomaly detection and prediction, IEEE Access 8 (2020) 45 184–45 197.
[23] F. Jauro, H. Chiroma, A.Y. Gital, M. Almutairi, S.M. Abdulhamid, J.H. Abawajy, Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend, Applied Soft Computing Journal 96 (2020), paper number 106582.
[24] J.-F. Chen, Q.H. Do, H.-N. Hsieh, Training artificial neural networks by a hybrid pso-cs algorithm, Algorithms 8 (2) (2015) 292–308.
[25] A.A. Heidari, H. Faris, S. Mirjalili, I. Aljarah, M. Mafarja, Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks, Nature-Inspired Optimizers (2020) 23–46.
[26] Y.-S. Park, S. Lek, Artificial neural networks: multilayer perceptron for ecological modeling, Developments in environmental modelling, vol. 28, ch. 7, Elsevier, 2016, pp. 123–140.
[27] R. Qaddoura, M. Al-Zoubi, H. Faris, I. Almomani, et al., A multi-layer classification approach for intrusion detection in iot networks based on deep learning, Sensors 21 (9) (2021) 2987.
[28] S. Zahara, S. Sugianto, Prediksi indeks harga konsumen komoditas makanan berbasis cloud computing menggunakan multilayer perceptron, JOINTECS (Journal of Information Technology and Computer Science) 6 (1) (2021) 21–28.
[29] G.S.B. Jahangeer, T.D. Rajkumar, Cloud storage based diagnosis of breast cancer using novel transfer learning with multi-layer perceptron, International Journal of System Assurance Engineering and Management (2022) 1–13.
[30] D. Mandic, J. Chambers, Recurrent neural networks for prediction: learning algorithms, architectures and stability, Wiley, 2001.
[31] Z.C. Lipton, J. Berkowitz, and C. Elkan, A critical review of recurrent neural
networks for sequence learning, arXiv preprint arXiv:1506.00019, 2015.
[32] R. Pascanu, T. Mikolov, and Y. Bengio, On the difficulty of training recurrent neural networks, in: Proceedings of International Conference on Machine Learning, 2013, pp. 1310–1318.
[33] P. Yazdanian, S. Sharifian, E2lg: a multiscale ensemble of lstm/gan deep learning architecture for multistep-ahead cloud workload prediction, The Journal of Supercomputing 77 (2021) 11 052–11 082.
[34] Y.S. Patel, J. Bedi, MAG-D: A multivariate attention network based approach for cloud workload forecasting, Future Generation Computer Systems (2023).
[35] S. Ouhame, Y. Hadi, A. Ullah, An efficient forecasting approach for resource utilization in cloud data center using cnn-lstm model, Neural Computing and Applications 33 (2021) 10 043–10 055.
[36] N. Tran, T. Nguyen, B.M. Nguyen, G. Nguyen, A multivariate fuzzy time series resource forecast model for clouds using lstm and data correlation analysis, Procedia Computer Science 126 (2018) 636–645.
[37] H.L. Leka, Z. Fengli, A.T. Kenea, A.T. Tegene, P. Atandoh, and N.W. Hundera, A hybrid cnn-lstm model for virtual machine workload forecasting in cloud data center, in: 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE, 2021, pp. 474–478.
[38] Z. Ding, J. Wang, Y. Cheng, and C. He, Alice: A lstm neural network based short-term power load forecasting approach in distributed cloud-edge environment, in: Journal of Physics: Conference Series, vol. 1624, no. 5. IOP Publishing, 2020, p. 052017.
[39] T. Hussain, K. Muhammad, W. Ding, J. Lloret, S.W. Baik, V.H.C. de Albuquerque, A comprehensive survey of multi-view video summarization, Pattern Recognition 109 (2021).
[40] K. Muhammad, T. Hussain, J. Del Ser, V. Palade, V.H.C. De Albuquerque, Deepres: A deep learning-based video summarization strategy for resource constrained industrial surveillance scenarios, IEEE Transactions on Industrial Informatics 16 (9) (2019) 5938–5947.
[41] K. Fukushima, S. Miyake, T. Ito, Neocognitron: A neural network model for a mechanism of visual pattern recognition, IEEE Transactions on Systems, Man and Cybernetics 13 (5) (1983) 826–834.
[42] A. Krizhevsky, I. Sutskever, and G. Hinton, Imagenet classification with deep convolutional neural network, in: Proceedings of the Conference of Neural Information Processing Systems, 2012, pp. 1106–1114.
[43] X. Zhang, X. Zhou, M. Lin, and J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of The IEEE Computer Vision and Pattern Recognition Conference, 2018.
[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in: Proceedings of The

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

IEEE Computer Vision and Pattern Recognition Conference, 2015.

[45] G. Huang, Z. Liu, L. van der Maaten, and K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of The IEEE Computer Vision and Pattern Recognition Conference, 2018.

[46] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Microsoft Research, Technical Report (2015) [Online]. Available:https://arxiv.org/pdf/1512.03385.pdf.

[47] A. Krizhevsky, I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of Conference on Neural Information Processing Systems. IEEE, 2012.

[48] K. Simonyan and A. Zisserman, Very deep convolutional networks for largescale image recognition, in: Proceedings of the International Conference on Learning Representations, 2015.

[49] A.D. Torres, H. Yan, A.H. Aboutalebi, A. Das, L. Duan, and P. Rad, Patient facial emotion recognition and sentiment analysis using secure cloud with hardware acceleration, in: Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications. Elsevier, 2018, pp. 61–89.

[50] A. Makkar, U. Ghosh, P.K. Sharma, Artificial intelligence and edge computing enabled web spam detection for next generation iot applications, IEEE Sensors Journal 21 (22) (2021) 25 352–25 361.

[51] D. Selvapandian, R. Santhosh, Deep learning approach for intrusion detection in iot-multi cloud environment, Automated Software Engineering 28 (2) (2021) 1–17.

[52] D. Scheinert and A. Acker, Telesto: A graph neural network model for anomaly classification in cloud services, in: Proceedings of International Conference on Service-Oriented Computing. Springer, 2020, pp. 214–227.

[53] W. Ullah, A. Ullah, I.U. Haq, K. Muhammad, M. Sajjad, S.W. Baik, CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks, Multimedia Tools and Applications 80 (11) (2021) 16 979–16 995.

[54] H. Saito, T. Aoki, K. Aoyama, Y. Kato, A. Tsuboi, A. Yamada, M. Fujishiro, S. Oka, S. Ishihara, T. Matsuda, et al., Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network, Gastrointestinal endoscopy 92 (1) (2020) 144–151.

[55] G. Shu, W. Liu, X. Zheng, J. Li, If-cnn: Image-aware inference framework for cnn with the collaboration of mobile devices and cloud, IEEE Access 6 (2018) 68 621–68 633.

[56] C.-H. Chen, C.-R. Lee, W.C.-H. Lu, Smart in-car camera system using mobile cloud computing framework for deep learning, Vehicular Communications 10 (2017) 84–90.

[57] B. Abu-Salih, M. Al-Tawil, I. Aljarah, H. Faris, P. Wongthongtham, K.Y. Chan, A. Beheshti, Relational learning analysis of social politics using knowledge graph embedding, Data Mining and Knowledge Discovery (2021) 1–40.

[58] P. Cui, X. Wang, J. Pei, W. Zhu, A survey on network embedding, IEEE Transactions on Knowledge and Data Engineering 31 (5) (2018) 833–852.

[59] W.L. Hamilton, R. Ying, and J. Leskovec, Representation learning on graphs: Methods and applications, in: Proceedings of IEEE Data Engineering Bulletin, 2017.

[60] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond euclidean data, IEEE Signal Processing Magazine 34 (4) (2017) 18–42.

[61] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, AI Open 1 (2020) 57–81.

[62] F. Frasca, E. Rossi, D. Eynard, B. Chamberlain, M. Bronstein, and F. Monti, Sign: Scalable inception graph neural networks, in: International Conference on Machine Learning, 2020.

[63] W. Hu, M. Shuaibi, A. Das, S. Goyal, A. Sriram, J. Leskovec, D. Parikh, and C.L. Zitnick, Forcenet: A graph neural network for large-scale quantum calculations, in: Proceedings of The Conference on Neural Information Processing Systems, 2021.

[64] J. Klicpera, S. Giri, J.T. Margraf, and S. Günnemann, Fast and uncertainty aware directional message passing for non-equilibrium molecules, in: Proceedings of The Conference on Neural Information Processing Systems, 2020.

[65] M. Shuaibi, A. Kolluru, A. Das, A. Grover, A. Sriram, Z. Ulissi, and C.L. Zitnick, Rotation invariant graph neural networks using spin convolutions, in: Proceedings of The Conference on Neural Information Processing Systems, 2021.

[66] J. Klicpera, F. Becker, and S. Günnemann, Gemnet: Universal directional graph neural networks for molecules, in: Proceedings of The Conference on Neural Information Processing Systems, 2021.

[67] Available: https://blog.twitter.com/engineering/en_us/topics/insights/2020/graph-ml-at-twitter.

[68] A. Sriram, A. Das, B.M. Wood, S. Goyal, and C.L. Zitnick, Towards training billion parameter graph neural networks for atomic simulations, in: Proceedings of The International Conference on Learning Representations, 2022.

[69] J. Gasteiger, F. Becker, and S. Gunnemann, Gemnet: Universal directional graph neural networks for molecules, in: Proceedings of The Conference on Neural Information Processing Systems, 2022.

[70] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S.Y. Philip, A comprehensive survey on graph neural networks, IEEE Transactions on Neural Networks and Learning Systems 32 (1) (2020) 4–24.

[71] R. Ying, R. He, K. Chen, P. Eksombatchai, W.L. Hamilton, and J. Leskovec, Graph convolutional neural networks for web-scale recommender systems, in: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018, pp. 974–983.

[72] L. Ge, Y. Jia, Q. Li, X. Ye, Dynamic multi-graph convolution recurrent neural network for traffic speed prediction, Journal of Intelligent & Fuzzy Systems no. Preprint (2023) 1–14.

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

[73] R. Yumlembam, B. Issac, S.M. Jacob, and L. Yang, Iot-based android malware detection using graph neural network with adversarial defense, IEEE Internet of Things Journal, 2022.
[74] Y. Li, S. Xie, Z. Wan, H. Lv, H. Song, Z. Lv, Graph-powered learning methods in the internet of things: A survey, Machine Learning with Applications 11 (2023).
[75] S. Wang, Y. Zheng, X. Jia, Secgnn: Privacy-preserving graph neural network training and inference as a cloud service, IEEE Transactions on Services Computing (2023).
[76] R. v. d. Berg, T.N. Kipf, and M. Welling, Graph convolutional matrix completion, arXiv preprint arXiv:1706.02263, 2017.
[77] S. Pratiher, S. Chattoraj, D. Nawn, M. Pal, R.R. Paul, H. Konik, and J. Chatterjee, A multi-scale context aggregation enriched mlp-mixer model for oral cancer screening from oral sub-epithelial connective tissues, in: 2022 30th European Signal Processing Conference (EUSIPCO). IEEE, 2022, pp. 1323–1327.
[78] Y. Li, K. Li, X. Liu, Y. Wang, L. Zhang, Lithium-ion battery capacity estimation–a pruned convolutional neural network approach assisted with transfer learning, Applied Energy 285 (2021).
[79] P.N. Suganthan, R. Katuwal, On the origins of randomization-based feedforward neural networks, Applied Soft Computing 105 (2021).
[80] T. Le-Cong, H.J. Kang, T.G. Nguyen, S.A. Haryono, D. Lo, X.-B.D. Le, and Q.T. Huynh, Autopruner: transformer-based call graph pruning, in: Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 2022, pp. 520–532.
[81] J. Poyatos, D. Molina, A.D. Martinez, J. Del Ser, F. Herrera, Evoprunedeeptl: An evolutionary pruning model for transfer learning based deep neural networks, Neural Networks 158 (2023) 59–82.
[82] D.T. Hoang, C. Lee, D. Niyato, P. Wang, A survey of mobile cloud computing: architecture, applications, and approaches, Wireless Communications and Mobile Computing 13 (2013) 1587–1611.
[83] Q. Zhang, L. Cheng, R. Boutaba, Cloud computing: State-of-the-art and research challenges, Journal of Internet Services and Applications 1 (1) (2010) 7–18.
[84] T. Dillon, C. Wu, and E. Chang, Cloud computing: Issues and challenges, in: Proceedings of the IEEE International Conference on Advanced Information Networking and Applications, 2010.
[85] J. Novet. (2021) How amazon's cloud business generates billions in profit. [Online]. Available: https://www.cnbc.com/2021/09/05/how-amazon-webservices-makes-money-estimated-margins-by-service.html.
[86] What is distributed cloud? 2022. https://www.windriver.com/solutions/learning/distributed-cloud.
[87] A. Smola and S. Narayanamurthy, An architecture for parallel topic models, Proceedings of the VLDB Endowment, vol. 3, no. 1–2, pp. 703–710, 2010.
[88] M. Li, D.G. Andersen, J.W. Park, A.J. Smola, A. Ahmed, V. Josifovski, J. Long, E.J. Shekita, and B.-Y. Su, Scaling distributed machine learning with the parameter server, in: Proceedings of Symposium on Operating Systems Design and Implementation, 2014, pp. 583–598.
[89] A. Harlap, A. Tumanov, A. Chung, G.R. Ganger, and P.B. Gibbons, Proteus: agile ML elasticity through tiered reliability in dynamic resource markets, in: Proceedings of the Twelfth European Conference on Computer Systems, 2017, pp. 589–604.
[90] A. Qiao, A. Aghayev, W. Yu, H. Chen, Q. Ho, G.A. Gibson, and E.P. Xing, Litz: Elastic framework for high-performance distributed machine learning, in: Proceedings of Annual Technical Conference, 2018, pp. 631–644.
[91] Y. Peng, Y. Bao, Y. Chen, C. Wu, C. Meng, W. Lin, Dl2: A deep learning-driven scheduler for deep learning clusters, IEEE Transactions on Parallel and Distributed Systems 32 (8) (2021) 1947–1960.
[92] Y. Chen, Y. Peng, Y. Bao, C. Wu, Y. Zhu, and C. Guo, Elastic parameter server load distribution in deep learning clusters, in: Proceedings of the 11th ACM Symposium on Cloud Computing, 2020, pp. 507–521.
[93] S. Wang, A. Pi, X. Zhou, Elastic parameter server: Accelerating ML training with scalable resource scheduling, IEEE Transactions on Parallel and Distributed Systems 33 (5) (2021) 1128–1143.
[94] L. Hu, J. Zhu, Z. Zhou, R. Cheng, X. Bai, and Y. Zhang, An optimal resource allocator of elastic training for deep learning jobs on cloud, arXiv preprint arXiv:2109.03389, 2021.
[95] T. Menouer, KCSS: kubernetes container scheduling strategy, The Journal of Supercomputing 77 (5) (2021) 4267–4293.
[96] N. Nakata, J.P. Chang, J.F. Lawrence, P. Boue, Body wave extraction and tomography at long beach california with ambient noise interferometry, Journal of Geophysical Research: Solid Earth 120 (2) (2015) 1159–1173.
[97] D.A. Philips, C.P. Santillan, L.M. Wang, R.W. King, W.M. Szeliga, T. Melbourne, M. Floyd, T.A. Herring, Plate boundary observatory and related networks: GPS data analysis methods and geodetic products, Reviews of Geophysics 54 (2016) 759–808.
[98] T. Cowles, J. Delaney, J. Orcutt, R. Weller, The ocean observatories initiative: sustained ocean observing across a range of spatial scales, Marine Technology Society Journal 44 (6) (2010) 54–64.
[99] I. Foster, D.B. Gannon, Cloud Computing For Science And Engineering, The MIT Press, 2017.
[100] T. Akidau, R. Bradshaw, C. Chambers, S. Chernyak, R.J. Fernandez-Moctezuma, R. Lax, S. McVeety, D. Mills, F. Perry, E. Schmidt, and S. Whittle, The dataflow model: a practical approach to balancing correctness, latency, and cost in massive scale, unbounded, out-of-order data processing, Proceedings of the VLDB ENdowment, vol. 8, no. 12, pp. 1792–1803, 2015.
[101] W. Rang, D.L. Yang, D.Z. Cheng, Y. Wang, Data life aware model updating strategy for stream-based online deep learning, IEEE Transactions on Parallel and Distributed Systems 32 (10) (2021) 2571–2581.
[102] A. Ashfahani, Autonomous deep learning: Incremental learning of deep neural networks for evolving data streams,

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

in: Proceedings of the International Conference on Data Mining Workshops, 2019, pp. 83–90.
[103] Y.L. Li, M. Zhang, W. Wang, Online real-time analysis of data streams based on an incremental high-order deep learning model, IEEE Access 6 (2018) 77 615–77 623.
[104] M. Pratama, W. Pedrycz, G.I. Webb, Online real-time analysis of data streams based on an incremental high-order deep learning model, IEEE Transactions on Fuzzy Systems 28 (7) (2020) 1315–1328.
[105] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet, A multi-task deep learning architecture for maritime surveillance using ais data streams, in: Proceedings of IEEE 5th International Conference on Data Science and Advanced Analytics, 2018, pp. 331–340.
[106] S.S. Zhang, J.W. Liu, and X. Zuo, Adaptive online incremental learning for evolving data streams, 2022. [Online]. Available: https://arxiv.org/abs/1805.04754.
[107] D.W. Otter, J.R. Medina, J.K. Kalita, A survey of the usages of deep learning for natural language processing, IEEE Transactions on Neural Networks and Learning Systems 32 (2) (2021) 604–624.
[108] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal 5 (4) (2014) 1093–1113.
[109] R. Feldman, Techniques and applications for sentiment analysis, Communications of the ACM 56 (4) (2013) 82–89.
[110] R. Obiedat, L. Al-Qaisi, R. Qaddoura, O. Harfoushi, A. Al-Zoubi, An intelligent hybrid sentiment analyzer for personal protective medical equipments based on word embedding technique: The COVID-19 era, Symmetry 13 (12) (2021) 2287.
[111] D.A. Al-Qudah, A.-Z. Ala'M, P.A. Castillo-Valdivieso, H. Faris, Sentiment analysis for e-payment service providers using evolutionary extreme gradient boosting, IEEE Access 8 (2020) 189 930–189 944.
[112] R. Obiedat, O. Harfoushi, R. Qaddoura, L. Al-Qaisi, A. Al-Zoubi, An evolutionary-based sentiment analysis approach for enhancing government decisions during COVID-19 pandemic: The case of jordan, Applied Sciences 11 (19) (2021) 9080.
[113] R.O. Sinnott and S. Cui, Benchmarking sentiment analysis approaches on the cloud, in: Proceedings of 2016 IEEE 22nd International Conference on Parallel and Distributed Systems, 2016, pp. 695–704.
[114] M. Ghorbani, M. Bahaghighat, Q. Xin, F. Özen, Convlstmconv network: a deep learning approach for sentiment analysis in cloud computing, Journal of Cloud Computing 9 (1) (2020) 1–12.
[115] M.R. Raza, W. Hussain, and J.M. Merigó, Long short-term memory-based sentiment classification of cloud dataset, in: Proceedings of IEEE Innovations in Intelligent Systems and Applications Conference, 2021, pp. 1–6.
[116] G. Preethi, P.V. Krishna, M.S. Obaidat, V. Saritha, and S. Yenduri, Application of deep learning to sentiment analysis for recommender system on cloud, in: Proceedings of International conference on computer, information and telecommunication systems. IEEE, 2017, pp. 93–97.
[117] M.A. Khan, S. Saqib, T. Alyas, A.U. Rehman, Y. Saeed, A. Zeb, M. Zareei, E.M. Mohamed, Effective demand forecasting model using business intelligence empowered with machine learning, IEEE Access 8 (2020) 116 013–116 023.
[118] B.M. Balachandran, S. Prasad, Challenges and benefits of deploying big data analytics in the cloud for business intelligence, Procedia Computer Science 112 (2017) 1112–1122.
[119] C. Moreno, R.A.C. González, E.H. Viedma, Data and artificial intelligence strategy: A conceptual enterprise big data cloud architecture to enable market-oriented organisations, International Journal of Interactive Multimedia and Artificial Intelligence 5 (6) (2019) 7–14.
[120] C. Juarez and H. Afli, Online news analysis on cloud computing platform for market prediction. in Collaborative European Research Conference, 2020, pp. 125–140.
[121] P. Dixit and S. Silakari, ''Deep learning algorithms for cybersecurity applications: A technological and status review," Computer Science Review, vol. 39, paper number 100317, 2021
[122] M.M. Alani, Big data in cybersecurity: a survey of applications and future trends, Journal of Reliable Intelligent Environments (2021) 1–30.
[123] P. Podder, S. Bharati, M. Mondal, P.K. Paul, and U. Kose, Artificial neural network for cybersecurity: a comprehensive review, arXiv preprint arXiv:2107.01185, 2021.
[124] L. Gupta, T. Salman, A. Ghubaish, D. Unal, A.K. Al-Ali, R. Jain, Cybersecurity of multi-cloud healthcare systems: A hierarchical deep learning approach, Applied Soft Computing 118 (2022), paper number 108439.
[125] Y. Chai, Y. Zhou, W. Li, Y. Jiang, An explainable multi-modal hierarchical attention model for developing phishing threat intelligence, IEEE Transactions on Dependable and Secure Computing 19 (2) (2022) 790–803.
[126] B. Abu-Salih, D.A. Qudah, M. Al-Hassan, S.M. Ghafari, T. Issa, I. Aljarah, A. Beheshti, and S. Alqahtan, An intelligent system for multi-topic social spam detection in microblogging, arXiv preprint arXiv:2201.05203, 2022.
[127] F.J. Abdullayeva, Advanced persistent threat attack detection method in cloud computing based on autoencoder and softmax regression algorithm, Array 10 (2021), paper number 100067.
[128] A. Vadariya and N.K. Jadav, A survey on phishing URL detection using artificial intelligence, in: Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications, 2021, pp. 9–20.
[129] L. Tang, Q.H. Mahmoud, A survey of machine learning-based solutions for phishing website detection, Machine Learning and Knowledge Extraction 3 (3) (2021) 672–694.
[130] G. Li, P. Sharma, L. Pan, S. Rajasegarar, C. Karmakar, N. Patterson, Deep learning algorithms for cyber security applications: A survey, Journal of Computer Security 29 (5) (2021) 447–471.
[131] K. Sethi, R. Kumar, N. Prajapati, and P. Bera, Deep reinforcement learning based intrusion detection system for

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

cloud infrastructure, in: Proceedings of the IEEE International Conference on Communication Systems and Networks, 2020, pp. 1–6.
[132] M.N. Hossain, J. Wang, O. Weisse, R. Sekar, D. Genkin, B. He, S.D. Stoller, G. Fang, F. Piessens, and E. Downing, Dependence-preserving data compaction for scalable forensic analysis, in: Proceedings of the 27th Security Symposium, 2018, pp. 1723–1740.
[133] M. Du, F. Li, G. Zheng, and V. Srikumar, Deeplog: Anomaly detection and diagnosis from system logs through deep learning, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1285–1298.
[134] T. Thilagam, R. Aruna, Intrusion detection for network based cloud computing by custom rc-nn and optimization, ICT Express 7 (4) (2021) 512–520.
[135] Y. Shen, E. Mariconti, P.A. Vervier, and G. Stringhini, Tiresias: Predicting security events through deep learning, in: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp.
592–605.
[136] M. Ramilli, Malware training sets: a machine learning dataset for everyone, Marco Ramilli Web Corner, 2016.
[137] N. Koroniotis, N. Moustafa, E. Sitnikova, B. Turnbull, Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset, Future Generation Computer Systems 100 (2019) 779–796.
[138] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, Hierarchical attention networks for document classification, in: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 2016, pp. 1480–1489.
[139] N. Moustafa and J. Slay, Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set), in: 2015 military communications and information systems conference (MilCIS). IEEE, 2015, pp. 1–6.
[140] L. Bergman and Y. Hoshen, Classification based anomaly detection for general data, in: Proceedings of International Conference on Learning Representations, 2020.
[141] Y. Wu, H.-N. Dai, and H. Tang, Graph neural networks for anomaly detection in industrial internet of things, IEEE Internet of Things Journal, 2021.
[142] H. Wang, Z. Wu, H. Jiang, Y. Huang, J. Wang, S. Kopru, and T. Xie, Groot: An event-graph-based approach for root cause analysis in industrial settings, in: Proceedings of IEEE/ACM International Conference on Automated Software Engineering, 2021.
[143] A. Protogerou, S. Papadopoulos, A. Drosou, D. Tzovaras, I. Refanidis, A graph neural network method for distributed anomaly detection in iot, Evolving Systems 12 (1) (2021) 19–36.
[144] F. Fusco, B. Eck, R. Gormally, M. Purcell, and S. Tirupathi, Knowledge-and data-driven services for energy systems using graph neural networks, in: Proceedings of IEEE International Conference on Big Data. IEEE, 2020, pp. 1301–1308.
[145] M. Lee, S. Hosseinalipour, C.G. Brinton, G. Yu, and H. Dai, A fast graph neural network-based method for winner determination in multi-unit combinatorial auctions, IEEE Transactions on Cloud Computing, 2020.
[146] M. Gao, Y. Li, and J. Yu, Workload prediction of cloud workflow based on graph neural network, in: Proceedings of International Conference on Web Information Systems and Applications, 2021, pp. 169–189.
[147] A. Rafiq, T.A. Khan, M. Afaq, and W.-C. Song, Service function chaining and traffic steering in sdn using graph neural network, in: Proceedings of IEEE International Conference on Information and Communication Technology Convergence, 2020, pp. 500–505.
[148] Z. Yu, W. Liu, X. Liu, and G. Wang, Drag-jdec: A deep reinforcement learning and graph neural network-based job dispatching model in edge computing, in: Proceedings of IEEE/ACM 29th International Symposium on Quality of Service. IEEE, 2021, pp. 1–10.
[149] T. Liang, X. Sheng, L. Zhou, Y. Li, H. Gao, Y. Yin, L. Chen, Mobile app recommendation via heterogeneous graph neural network in edge computing, Applied Soft Computing vol. 103, paper number 107162 (2021).
[150] A. Said, S.-U. Hassan, S. Tuarob, R. Nawaz, M. Shabbir, DGSD: Distributed graph representation via graph statistical properties, Future Generation Computer Systems 119 (2021) 166–175.
[151] S. Chen, J. Dong, P. Ha, Y. Li, S. Labi, Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles, Computer-Aided Civil and Infrastructure Engineering 36 (7) (2021) 838–857.
[152] Y. Gao, X. Gu, H. Zhang, H. Lin, and M. Yang, ''Runtime performance prediction for deep learning models with graph neural network,'' Technical Report MSR-TR-2021-3. Microsoft, Tech. Rep., 2021.
[153] W. Ullah, A. Ullah, T. Hussain, Z.A. Khan, S.W. Baik, An efficient anomaly recognition framework using an attention residual lstm in surveillance videos, Sensors 21 (8) (2021) 2811.
[154] W. Ullah, A. Ullah, T. Hussain, K. Muhammad, A.A. Heidari, J. Del Ser, S.W. Baik, V.H.C. De Albuquerque, Artificial intelligence of things-assisted two stream neural network for anomaly detection in surveillance big video data, Future Generation Computer Systems 129 (2022) 286–297.
[155] W. Xu, L. Huang, A. Fox, D. Patterson, and M. Jordan, ''Online system problem detection by mining patterns of console logs,'' in: 2009 ninth IEEE international conference on data mining. IEEE, 2009, pp. 588–597.
[156] D. Duplyakin, R. Ricci, A. Maricq, G. Wong, J. Duerig, E. Eide, L. Stoller, M. Hibler, D. Johnson, K. Webb et al., The design and operation of fCloudLabg", in: 2019 USENIX annual technical conference (USENIX ATC 19), 2019, pp. 1–14.
[157] X. Zhou, C. Xu, B. Kimmons, Detecting tourism destinations using scalable geospatial analysis based on

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License. http://creativecommons.org/licenses/by/4.0/

cloud computing platform, Computers, Environment and Urban Systems 54 (2015) 144–153.
[158] K. Vassakis, E. Petrakis, I. Kopanakis, J. Makridis, and G. Mastorakis, Locationbased social network data for tourism destinations, in: Proceedings of Big data and innovation in tourism, travel, and hospitality, 2019, pp. 105–114.
[159] W. Wang, N. Kumar, J. Chen, Z. Gong, X. Kong, W. Wei, H. Gao, Realizing the potential of the internet of things for smart tourism with 5g and ai, IEEE Network 34 (6) (2020) 295–301.
[160] S.Y. Park, B. Pan, Identifying the next non-stop flying market with a big data approach, Tourism Management 66 (2018) 411–421.
[161] A. Kontogianni, E. Alepis, C. Patsakis, Promoting smart tourism personalized services via a combination of deep learning techniques, Expert Systems with Applications vol. 187, paper number 115964 (2022).
[162] M. Mariani, R. Baggio, Big data and analytics in hospitality and tourism: a systematic literature review, International Journal of Contemporary Hospitality Management no. 1, paper number 232 (2021).
[163] J.C. Cepeda-Pacheco, M.C. Domingo, Deep learning and internet of things for tourist attraction recommendations in smart cities, Neural Computing and Applications (2022) 1–19.
[164] J. Guerra-Montenegro, J. Sanchez-Medina, I. La na, D. Sanchez-Rodriguez, I. Alonso-Gonzalez, and J. Del Ser, ‘‘Computational intelligence in the hospitality industry: A systematic literature review and a prospect of challenges,” Applied Soft Computing, vol. 102, p. 107082, 2021.
[165] F. Piccialli, F. Giampaolo, G. Casolla, V.S. Di Cola, K. Li, A deep learning approach for path prediction in a location-based iot system, Pervasive and Mobile Computing vol. 66, paper number 101210 (2020).
[166] Y.-C. Chang, C.-H. Ku, C.-H. Chen, Using deep learning and visual analytics to explore hotel reviews and responses, Tourism Management 80 (2020).
[167] G. Díaz, H. Macià, V. Valero, J. Boubeta-Puig, F. Cuartero, An intelligent transportation system to control air pollution and road traffic in cities integrating cep and colored petri nets, Neural Computing and Applications 32 (2) (2020) 405–426.
[168] P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, A. Mouzakitis, A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles, IEEE Transactions on Intelligent Transportation Systems (2021).
[169] Q. Cai, M. Abdel-Aty, Y. Sun, J. Lee, J. Yuan, Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data, Transportation Research Part A: Policy and Practice 127 (2019) 71–85.
[170] J. Wang, R. Chen, Z. He, Traffic speed prediction for urban transportation network: A path based deep learning approach, Transportation Research Part C: Emerging Technologies 100 (2019) 372–385.
[171] X. Wang, J. Liu, T. Qiu, C. Mu, C. Chen, P. Zhou, A real-time collision prediction mechanism with deep learning for intelligent transportation system, IEEE Transactions on Vehicular Technology 69 (9) (2020) 9497–9508.
[172] S. Yang, W. Ma, X. Pi, S. Qian, A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources, Transportation Research Part C: Emerging Technologies 107 (2019) 248–265.
[173] A. Paranjothi, M.S. Khan, S. Zeadally, A survey on congestion detection and control in connected vehicles, Ad Hoc Networks 108 (2020).
[174] C. Chen, B. Liu, S. Wan, P. Qiao, Q. Pei, An edge traffic flow detection scheme based on deep learning in an intelligent transportation system, IEEE Transactions on Intelligent Transportation Systems 22 (3) (2020) 1840–1852.
[175] M.M. Iqbal, M.T. Mehmood, S. Jabbar, S. Khalid, A. Ahmad, G. Jeon, An enhanced framework for multimedia data: Green transmission and portrayal for smart traffic system, Computers & Electrical Engineering 67 (2018) 291–308.
[176] R.S. Thejaswini and S. Rajaraajeswari, A real-time traffic congestion avoidance framework for smarter cities, in: Proceedings of AIP Conference Proceedings, vol. 2039, no. 1, paper number 020009, 2018.
[177] S. Sarker, B. Wankum, T. Perey, M.M. Mau, J. Shimizu, R. Jones, B.S. Terry, A novel capsule-delivered enteric drug-injection device for delivery of systemic biologics: A pilot study in a porcine model, IEEE Transactions on Biomedical Engineering 69 (6) (2022) 1870–1879.
[178] S. Sangodoyin, E.M. Ugurlu, M. Dey, M. Prvulovic, A. Zajic´, Leveraging on-chip transistor switching for communication and sensing in neural implants and gastrointestinal devices, IEEE Transactions on Biomedical Engineering 69 (1) (2021) 377–389.
[179] S. Li, P. Si, Z. Zhang, J. Zhu, X. He, N. Zhang, Dfca-net: Dual feature context aggregation network for bleeding areas segmentation in wireless capsule endoscopy images, Journal of Medical and Biological Engineering 42 (2) (2022) 179–188.
[180] M.R. Basar, F. Malek, K.M. Juni, M.S. Idris, M.I.M. Saleh, Ingestible wireless capsule technology: A review of development and future indication, International Journal of Antennas and Propagation vol. 2012, paper number 807165 (2012).
[181] I. Mehmood, M. Sajjad, S.W. Baik, Mobile-cloud assisted video summarization framework for efficient management of remote sensing data generated by wireless capsule sensors, Sensors 14 (9) (2014) 17 112–17 145.
[182] R. Hamza, K. Muhammad, Z. Lv, F. Titouna, Secure video summarization framework for personalized wireless capsule endoscopy, Pervasive and Mobile Computing 41 (2017) 436–450.
[183] K. Muhammad, S. Khan, N. Kumar, J. Del Ser, S. Mirjalili, Vision-based personalized wireless capsule endoscopy for smart healthcare: Taxonomy, literature review, opportunities and challenges, Future Generation Computer Systems 113 (2020) 266–280.

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

[184] K. Muhammad, M. Sajjad, M.Y. Lee, S.W. Baik, Efficient visual attention driven framework for key frames extraction from hysteroscopy videos, Biomedical Signal Processing and Control 33 (2017) 161–168.
[185] K. Muhammad, J. Ahmad, M. Sajjad, S.W. Baik, Visual saliency models for summarization of diagnostic hysteroscopy videos in healthcare systems, SpringerPlus vol. 5, paper number 1495 (2016).
[186] R. Shrestha, S.K. Mohammed, M.M. Hasan, X. Zhang, K.A. Wahid, Automated adaptive brightness in wireless capsule endoscopy using image segmentation and sigmoid function, IEEE Transactions on Biomedical Circuits and Systems 10 (4) (2016) 884–892.
[187] J.-Y. He, X. Wu, Y.-G. Jiang, Q. Peng, R. Jain, Hookworm detection in wireless capsule endoscopy images with deep learning, IEEE Transactions on Image Processing 27 (5) (2018) 2379–2392.
[188] X. Wu, H. Chen, T. Gan, J. Chen, C.-W. Ngo, Q. Peng, Automatic hookworm detection in wireless capsule endoscopy images, IEEE Transactions on Medical Imaging 35 (7) (2016) 1741–1752.
[189] Y. Yuan, J. Wang, B. Li, M.Q.-H. Meng, Saliency based ulcer detection for wireless capsule endoscopy diagnosis, IEEE Transactions on Medical Imaging 34 (10) (2015) 2046–2057.
[190] I. Mehmood, M. Sajjad, S.W. Baik, Video summarization based tele endoscopy: a service to efficiently manage visual data generated during wireless capsule endoscopy procedure, Journal of medical systems 38 (9) (2014) 1–9.
[191] Y. Yuan, M.Q.-H. Meng, Deep learning for polyp recognition in wireless capsule endoscopy images, Medical physics 44 (4) (2017) 1379–1389.
[192] X. Jia, M.Q.-H. Meng, A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images, 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE 2016 (2016) 639–642.
[193] S. Jain, A. Seal, A. Ojha, A. Yazidi, J. Bures, I. Tacheci, O. Krejcar, A deep cnn model for anomaly detection and localization in wireless capsule endoscopy images, Computers in Biology and Medicine 137 (2021).
[194] T. Aoki, A. Yamada, K. Aoyama, H. Saito, A. Tsuboi, A. Nakada, R. Niikura, M. Fujishiro, S. Oka, S. Ishihara, et al., Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network, Gastrointestinal endoscopy 89 (2) (2019) 357–363.
[195] X. Gu, K. Nahrstedt, A. Messer, I. Greenberg, D. Milojicic, Adaptive offloading for pervasive computing, IEEE Pervasive Computing 3 (3) (2004) 66–73.
[196] K. Yang, S. Ou, H.-H. Chen, On effective offloading services for resource constrained mobile devices running heavier mobile internet applications, IEEE Communications Magazine 46 (1) (2008) 56–63.
[197] A.P. Miettinen and J.K. Nurminen, Energy efficiency of mobile clients in cloud computing, in: Proceedings of 2nd USENIX Workshop on Hot Topics in Cloud Computing, 2010.
[198] J.-C. Hsieh, A.-H. Li, C.-C. Yang, Mobile, cloud, and big data computing: contributions, challenges, and new directions in tele cardiology, International Journal of Environmental Research and Public Health 10 (11) (2013) 6131–6153.
[199] G. Fortino, D. Parisi, V. Pirrone, G. Di Fatta, Bodycloud: A SaaS approach for community body sensor networks, Future Generation Computer Systems 35 (2014) 62–79.
[200] C. Xia, J. Zhao, H. Cui, X. Feng, J. Xue, Dnntune: Automatic benchmarking dnn models for mobile-cloud computing, ACM Transactions on Architecture and Code Optimization (TACO) 16 (4) (2019) 1–26.
[201] B. Kumar, G. Pandey, B. Lohani, S.C. Misra, A multi-faceted cnn architecture for automatic classification of mobile lidar data and an algorithm to reproduce point cloud samples for enhanced training, ISPRS journal of photogrammetry and remote sensing 147 (2019) 80–89.
[202] A.E. Eshratifar, M.S. Abrishami, M. Pedram, JointDNN: An efficient training and inference engine for intelligent mobile cloud computing services, IEEE Transactions on Mobile Computing 20 (2) (2019) 565–576.
[203] A. Katal, S. Dahiya, T. Choudhury, A survey on cloud computing in energy management of the smart grids, Cluster Computing (2021).
[204] W. Pedrycz, Towards green machine learning: challenges, opportunities, and developments, Journal of Smart Environments and Green Computing 2 (4) (2022) 163–174.
[205] R. Schwartz, J. Dodge, N.A. Smith, O. Etzioni, Green ai, Communications of the ACM 63 (12) (2020) 54–63.
[206] H. Jin, Q. Song, and X. Hu, Auto-keras: An efficient neural architecture search system, in: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2019, pp. 1946–1956.
[207] A. Li, O. Spyra, S. Perel, V. Dalibard, M. Jaderberg, C.J. Gu, D. Budden, T. Harley, and P. Gupta, A generalized framework for population based training, in: Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2019.
[208] T. Elsken, J.H. Metzen, F. Hutter, Neural architecture search: A survey, The Journal of Machine Learning Research 20 (1) (2019) 1997–2017.
[209] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, X. Chen, X. Wang, A comprehensive survey of neural architecture search: Challenges and solutions, ACM Computing Surveys (CSUR) 54 (4) (2021) 1–34.
[210] H. Li, P. Wu, Z. Wang, J. f. Mao, F.E. Alsaadi, and Z.N. Y, ‘‘A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer,” Computers in Biology and Medicine, vol. 151A, 2022, 106265.
[211] H. Li, N.Y. Zeng, P.S. Wu, K. Clawson, Cov-net: A computer-aided diagnosis method for recognizing covid-19 from chest x-ray images via machine vision, Expert Systems with Applications 207 (2022).
[212] P. Wu, Z. Wang, B. Zheng, H. Li, F.E. Alsaadi, N. Zeng, Aggn: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion, Computers in Biology and Medicine 152 (2023).

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

[213] I. Makdoom, M. Abolhasan, H. Abbas, W. No, Blockchains adoption in iot: the challenges, and a way forward, Journal of Network and Computer Applications 125 (2019) 251–279.
[214] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P.S. Yu, X. Zhang, Membership inference attacks on machine learning: A survey, ACM Computing Surveys (CSUR) 54 (11s) (2022) 1–37.
[215] M. Jegorova, C. Kaul, C. Mayor, A.Q. O'Neil, A. Weir, R. Murray-Smith, S.A. Tsaftaris, Survey: Leakage and privacy at inference time, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).
[216] M. Abadi, A. Chu, I. Goodfellow, H.B. McMahan, I. Mironov, K. Talwar, and L. Zhang, Deep learning with differential privacy, in: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, 2016, pp. 308–318.
[217] C. Dwork, ''Differential privacy: A survey of results,'' in Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5. Springer, 2008, pp. 1–19.
[218] S. Meftah, B.H.M. Tan, C.F. Mun, K.M.M. Aung, B. Veeravalli, V. Chandrasekhar, Doren: toward efficient deep convolutional neural networks with fully homomorphic encryption, IEEE Transactions on Information Forensics and Security 16 (2021) 3740–3752.
[219] Z. Zhong, W. Bao, J. Wang, X. Zhu, X. Zhang, Flee: A hierarchical federated learning framework for distributed deep neural network over cloud, edge, and end device, ACM Transactions on Intelligent Systems and Technology (TIST) 13 (5) (2022) 1–24.
[220] R. Gupta, I. Gupta, D. Saxena, A.K. Singh, A differential approach and deep neural network based data privacy-preserving model in cloud environment, Journal of Ambient Intelligence and Humanized Computing (2022) 1–16.
[221] J. Vasa, A. Thakkar, Deep learning: Differential privacy preservation in the era of big data, Journal of Computer Information Systems (2022) 1–24.
[222] M.H. Rahman, M.M. Mowla, S. Shanto, Differential privacy enabled deep neural networks for wireless resource management, Mobile Networks and Applications 27 (5) (2022) 2153–2162.
[223] J. Xiong, H. Zhu, Real-time trajectory privacy protection based on improved differential privacy method and deep learning model, Journal of Cloud Computing 11 (1) (2022) 1–15.
[224] F. Gava, L.M. Bayati, A scalable algorithm for homomorphic computing on multi-core clusters, 2022 21st International Symposium on Parallel and Distributed Computing (ISPDC), IEEE (2022) 57–64.
[225] S. Meftah, B.H.M. Tan, K.M.M. Aung, L. Yuxiao, L. Jie, B. Veeravalli, Towards high performance homomorphic encryption for inference tasks on cpu: An mpi approach, Future Generation Computer Systems 134 (2022) 13–21.
[226] J.A. Alzubi, O.A. Alzubi, A. Singh, M. Ramachandran, Cloud-iiot-based electronic health record privacy-preserving by cnn and blockchain-enabled federated learning, IEEE Transactions on Industrial Informatics 19 (1) (2022) 1080–1087.
[227] Z. Liu, Z. Gao, J. Wang, Q. Liu, J. Wei, et al., Ppefl: An edge federated learning architecture with privacy-preserving mechanism, Wireless Communications and Mobile Computing 2022 (2022).
[228] J. Fiosina, ''Interpretable privacy-preserving collaborative deep learning for taxi trip duration forecasting,'' in Smart Cities, Green Technologies, and Intelligent Transport Systems: 10th International Conference, SMARTGREENS 2021, and 7th International Conference, VEHITS 2021, Virtual Event, April 28–30, 2021, Revised Selected Papers. Springer, 2022, pp. 392–411.
[229] B. Jan, H. Farman, M. Khan, M. Imran, I.U. Islam, A. Ahmad, S. Ali, G. Jeon, Deep learning in big data analytics: a comparative study, Computers & Electrical Engineering 75 (2019) 275–287.
[230] J. Dean, G.S. Corrado, R. Monga, K. Chen, M. Devin, Q.V. Le, M.Z. Mao, M. Ranzato, A. Senior, P. Tucker et al., Large scale distributed deep networks, in: Proceedings of Advances in Neural Information Processing Systems, 2012.
[231] B. Abu-Salih, P. Wongthongtham, D. Zhu, K.Y. Chan, A. Rudra, Social Big Data Analytics: Practices, Techniques, and Applications, Springer Nature, 2021.
[232] A.A. Taha, W. Ramo, H.H.K. Alkhaffaf, Impact of external auditor–cloud specialist engagement on cloud auditing challenges, Journal of Accounting & Organizational Change 17 (3) (2021) 309–331.
[233] Z. Zhang, C. Wu, D.W. Cheung, A survey on cloud interoperability: taxonomies, standards, and practice, ACM SIGMETRICS Performance Evaluation Review 40 (4) (2013) 13–22.
[234] H. Schulze, Cloud security report, Fortinet, Report, 2021. [Online]. Available: https://www.fortinet.com/content/dam/fortinet/assets/analyst-reports/arcybersecurity-cloud-security.pdf.
[235] C. Ramalingam, P. Mohan, Addressing semantics standards for cloud portability and interoperability in multi cloud environment, Symmetry 13 (2) (2021) 317.
[236] Available: https://insightaas.com/cloud-interoperability-andportability-necessary-or-nice-to-have/.
[237] R. Ranjan, The cloud interoperability challenge, IEEE Cloud Computing 1 (2) (2014) 20–24.
[238] A. Romasanta and J. Wareham, Fair data through a federated cloud infrastructure: Exploring the science mesh, in: Proceedings of European Conference on Information Systems, 2021.
[239] B. Abu-Salih, Domain-specific knowledge graphs: A survey, Journal of Network and Computer Applications 185 (2021).
[240] H. Mezni, M. Sellami, S. Aridhi, F.B. Charrada, Towards big services: a synergy between service computing and parallel programming, Computing 103 (11) (2021) 2479–2519.
[241] O. Adedugbe, E. Benkhelifa, R. Campion, F. Al-Obeidat, A.B. Hani, U. Jayawickrama, Leveraging cloud computing for the semantic web: review and trends, Soft Computing 24 (8) (2020) 5999–6014.

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

[242] I. Grangel-González, F. Lösch, and A. ul Mehdi, Knowledge graphs for efficient integration and access of manufacturing data, in: Proceedings of 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), vol. 1. IEEE, 2020, pp. 93–100.

[243] G. Ditzler, M. Roveri, C. Alippi, R. Polikar, Learning in nonstationary environments: A survey, IEEE Computational Intelligence Magazine 10 (4) (2015) 12–25.

[244] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, Learning under concept drift: A review, IEEE transactions on knowledge and data engineering 31 (12) (2018) 2346–2363.

[245] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, T. Tuytelaars, A continual learning survey: Defying forgetting in classification tasks, IEEE transactions on pattern analysis and machine intelligence 44 (7) (2021) 3366–3385.

[246] T. Khan, W. Tian, G. Zhou, S. Ilager, M. Gong, R. Buyya, Machine learning (ml)–centric resource management in cloud computing: A review and future directions, Journal of Network and Computer Applications (2022).

[247] R.S. Sutton, A.G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[248] V. François-Lavet, P. Henderson, R. Islam, M.G. Bellemare, and J. Pineau, An introduction to deep reinforcement learning, arXiv preprint arXiv:1811.12560, 2018.

[249] Y. Li, Deep reinforcement learning: An overview, arXiv preprint arXiv:1701.07274, 2017.

[250] B.R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A.A. Al Sallab, S. Yogamani, P. Pérez, Deep reinforcement learning for autonomous driving: A survey, IEEE Transactions on Intelligent Transportation Systems (2021) 4909–4926.

[251] F. Zeng, C. Wang, and S.S. Ge, ‘‘A survey on visual navigation for artificial agents with deep reinforcement learning,’’ IEEE Access, vol. 8, pp. 135 426–135 442, 2020.

[252] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, Deep reinforcement learning: A brief survey, IEEE Signal Processing Magazine 34 (6) (2017) 26–38.

[253] K. Shao, Z. Tang, Y. Zhu, N. Li, and D. Zhao, A survey of deep reinforcement learning in video games, arXiv preprint arXiv:1912.10944, 2019.

[254] A. Gupta, A. Anpalagan, L. Guan, A.S. Khwaja, Deep learning for object detection and scene perception in self-driving cars:survey, challenges, and open issues, Array vol. 10, paper number 100057 (2021).

[255] N.Y. Zeng, P.S. Wu, Z.D. Wang, H. Li, W.B. Liu, X.H. Liu, A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection, IEEE Transactions on Instrumentation and Measurement 71 (2022).

[256] N. Rodríguez-Barroso, A.R. Moya, J.A. Fernández, E. Romero, E. Martínez-Cámara, and F. Herrera, Deep learning hyper-parameter tuning for sentiment analysis in twitter based on evolutionary algorithms, in: 2019 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2019, pp. 255–264.

[257] S.R. Young, D.C. Rose, T.P. Karnowski, S.-H. Lim, and R.M. Patton, Optimizing deep learning hyper-parameters through an evolutionary algorithm, in: Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments, 2015, pp. 1–5.

[258] R. Tanabe and A. Fukunaga, Success-history based parameter adaptation for differential evolution, in: Proceedings of IEEE Congress on Evolutionary Computation, 2013, pp. 71–78.

[259] M. Wu, W. Su, L. Chen, Z. Liu, W. Cao, K. Hirota, Weightadapted convolution neural network for facial expression recognition in human-robot interaction, IEEE Transactions on Systems, Man, and Cybernetics - Systems 51 (3) (2021) 1473–1484.

[260] M. Gong, J. Liu, A. Qin, K. Zhao, K.C. Tan, Evolving deep neural networks via cooperative coevolution with backpropagation, IEEE Transactions on Neural Networks and Learning Systems 32 (1) (2021) 420–434.

[261] M.R. Chen, B.P. Chen, G.Q. Zeng, K. Lu, P. Chu, An adaptive fractional-order bp neural network based on extremal optimization for handwritten digits recognition, Neuro computing 391 (2020) 260–272.

[262] Z. Lu, G. Sreekumar, E. Goodman, W. Banzhaf, K. Deb, V.N. Boddeti, Neural architecture transfer, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2021).

[263] B. Ma, X. Li, Y. Xia, Y. Zhang, Autonomous deep learning: A genetic dcnn designer for image classification, Neurocomputing 379 (2020) 152–161.

[264] F. Mattioli, D. Caetano, A. Cardoso, E. Naves, and E. Lamounier, An experiment on the use of genetic algorithms for topology selection in deep learning, Journal of Electrical and Computer Engineering, no. Article ID 3217542, 2019.

[265] B. Wang, Y. Sun, B. Xue, and M. Zhang, A hybrid ga-pso method for evolving architecture and short connections of deep convolutional neural networks, in: Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence, 2019.

[266] M. Suganuma, S. Shirakawa, and T. Nagao, A genetic programming approach to designing convolutional neural network architectures, in: Proceedings of Twenty-Seventh International Joint Conference on Artificial Intelligence, 2017, pp. 5369–5373.

[267] A.D. Martinez, J. Del Ser, E. Villar-Rodriguez, E. Osaba, J. Poyatos, S. Tabik, D. Molina, F. Herrera, Lights and shadows in evolutionary deep learning: Taxonomy, critical methodological analysis, cases of study, learned lessons, recommendations and challenges, Information Fusion 67 (2021) 161–194.

[268] T. Serizawa and H. Fujita. (2021) Optimization of convolutional neural network using the linearly decreasing weight particle swarm optimization. [Online]. Available: https://arxiv.org/abs/2001.05670.

[269] J. Fregoso, C.I. Gonzalez, G.E. Martinez, Optimization of convolutional neural networks architectures using pso for sign language recognition, Axioms vol. 10, number paper 139 (2021).

[270] Z. Fouad, M. Alfonse, M. Roushdy, A.-B.M. Salem, Hyper-parameter optimization of convolutional neural network based on particle swarm optimization algorithm, Bulletin of Electrical Engineering and Informatics 10 (6) (2021) pp.

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/

[271] X. Chen, Y. Sun, M. Zhang, D. Peng, Evolving deep convolutional variational autoencoders for image classification, IEEE Transactions on Evolutionary Computation 25 (5) (2021).
[272] T. Zhang, C. Lei, Z. Zhang, X.-B. Meng, C.L.P. Chen, As-nas: Adaptive scalable neural architecture search with reinforced evolutionary algorithm for deep learning, IEEE Transactions on Evolutionary Computation 25 (5) (2021).
[273] L. Xie and A. Yuille, Genetic cnn, in: Proceedings of IEEE International Conference on Computer Vision, 2017.
[274] S. Pimminger, S. Wagner, W. Kurschl, and J. Heinzelreiter, Optimization as a service: On the use of cloud computing for metaheuristic optimization, in: Proceedings of the International Conference on Computer Aided Systems Theory, 2013, p. 348–355.
[275] A.D. Martinez, J.D. Ser, E. Villar-Rodriguez, E. Osaba, J. Poyatos, S. Tabik, D. Molina, F. Herrera, Lights and shadows in evolutionary deep learning taxonomy, critical methodological analysis, cases of study, learned lessons, recommendations and challenges, Information Fusion 67 (2021) 161–194.

©2023 The Author(s). Published by Infogain Publication, This work is licensed under a Creative Commons Attribution 4.0 License.
http://creativecommons.org/licenses/by/4.0/